Green GDP: The key to sustainable development

Xiao Liu [1, *], Penghao Chen[1], Shaohua Wen[1]

[1] School of Computer Science and Software Engineering, University of Science and Technology Liaoning, Anshan, China

* Corresponding Author Email: 3151184496@qq.com

Abstract. Green Gross Domestic Product (GGDP) is an economic indicator designed to measure the environmental losses caused by economic growth. It takes into account factors such as the value of ecosystem services, environmental pollution and depletion of natural resources, thus providing a more comprehensive way of assessing economic development. Green growth is a special strategy for sustainable development, which provides a way to solve environmental problems and use natural resources.First, we investigate the impact of green technologies and growth on environmental factors in gross domestic product (GDP) in multiple countries from 2000 to 2020. In addition, it explores the linear and nonlinear effects of green GDP growth. The results show that the linear effect of GGDP has a positive effect on green growth. Conversely, the nonlinear effects of GDP have a negative impact on green growth.We then used the Bayer-Hanch cointegration test on a sample of G7 countries to study the impact of technological innovation (green patents) on green growth (e.g. EAMFP). Through indicators such as green growth, GDP, energy consumption, green technology innovation and pollution emissions, a multivariate linear model is established to estimate its impact on global climate change mitigation. Finally, we discuss the advantages and disadvantages of the model, leaving room for further study.

Keywords: green GDP, greenhouse effect, sustainable development.

1. Introduction

In the world, gross domestic product (GDP) is usually used as an indicator to measure the economic development level of a country or region in a certain period of time, which is widely used in macroeconomic evaluation. However, GDP only focuses on economic benefits without considering the environmental cost of economic growth, and cannot quantitatively reveal the contribution of ecosystem to economic growth. Objectively speaking, GDP accounting is carried out at the cost of ecological environmental damage and resource consumption.

2. GGDP accounting system

2.1. Data preprocessing

First of all, the data obtained by the search for cleaning, due to the data in the process of recording there will be errors, these errors may come from the artificial omission, or the failure of the machine record, first of all, we give the data for outlier processing, using the Laida criterion.Let the measured data be measured with equal precision，Independently obtained $x_1,x_2,...,x_n$,Calculate the arithmetic mean x and the residual error $v_i=x_i-x(i=1,2,...,n)$, And calculate the standard deviation by Bessel formula δ, if a residual error of measured values x_b $v_b(1\leq b\leq n)$, Satisfies the following equation：

$$|v_b| = |x_b - x| > 3\delta \tag{1}$$

We consider x_b as a bad value containing coarse error values and should be Eliminated.

2.2. Using mathematical models to process data

Green GDP is mainly a measurement standard system, and we can build statistical models and use hypothesis testing methods to solve them.

The t-test is also used to determine the significant difference between the sample mean and the population mean. In many ways a t-test is similar to a z-test, but the biggest difference is that the theoretical variance of the population is unknown, and the t-distribution can only be estimated from sample data. The independent sample t-test was used to analyze the difference between the categorical data and the quantitative data, and the paired sample t-test was used to reveal the contrast relationship between the quantitative data. The order of the samples should correspond to one another. The statistic for the paired t-test is defined as follows:

$$t = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (2)$$

The null hypothesis of the t-test is that the statistic is not significantly different between the two samples, and the alternative hypothesis is that the statistic is significantly different between the two samples. Usually, if the probability is less than 0.05, the alternative hypothesis is accepted with a significance of 0.05.

2.3. Solving the model

According to Equations (1) and (2), we used paired sample t-test for continuous attributes in the data set, chi-square independence test for discrete attributes, and took brand as an independent feature. The statistical test results are shown in Table 1 and Table 2:

Table 1. Results of paired samples t-test

Month	Average value1	Average value2	Test statistic	Probability value	Significance degree 0.05
1	87.22872288	77.42985004	11.00384105	2.27E-27	Yes
2	86.52311285	77.68413748	9.681982855	1.09E-21	Yes
3	85.819737	75.37141921	9.927481645	1.07E-22	Yes
4	86.64555328	78.42743528	8.885136976	1.41E-18	Yes
5	85.60489742	76.70843588	9..29654399	3.74E-20	Yes
6	85.3760682	77.4811245	8.309990915	1.76E-16	Yes
7	85.55914104	77.62103504	8.482236407	4.28E-17	Yes
8	84.92170592	77.19093868	7.940063397	3.37E-15	Yes
9	24.57575758	21.21394102	2.853356466	0.004371374	Yes
10	7.96969697	7.618766756	0.824196669	0.409928	No
11	3.464646465	3.455227882	0.084470731	0.932690791	No
12	0.888888889	0.866487936	0.35915445	0.719518197	No
1	34.08080808	34.31581769	-0.455193771	0.649020207	No
2	9.666666667	10.10348525	-0.863150273	0.388160347	No
3	31.27272727	26.52915335	3.644532214	0.000274845	Yes
4	19.17171717	16.44718499	2.42519336	0.015389957	Yes
5	20.83838384	16.49302949	4.509567429	6.88E-06	Yes
6	3.717171717	15.89383378	-9.174430345	1.12E-19	Yes
7	2.292929293	9.978552279	-6.040665	1.83E-09	Yes

Table 2. Chi-square test of independence results

Month	Test statistic	Probability value	0.05 significance degree	0.01 significance degree	0.001 significance degree
2018-1	2.178295885	0.902592	No	No	No
2005-3	4.79505183	0.252836333	No	No	No
2016-6	5.807122075	0.99004826	No	No	No
2009-9	7.421281106	0.68516416	No	No	No
2013-1	8.59782964	0.416971242	No	No	No
2009-2	9.52358344	0.612813496	No	No	No
Difference	36.8402086	0.00719347086	Yes	Yes	Yes

2.4. GGDP accounting method

The accounting theory of green GDP is developing continuously, and the content contained in the accounting system is constantly enriched. Although there are still many accounting systems and the accounting content is also different, However, the system and contents basically cover three aspects: the cost of resource consumption reduction including water, energy and cultivated land resources, the cost of environmental degradation including the cost of environmental pollution consumption reduction and environmental protection expenditure, and the benefit of resource and environment improvement. In terms of empirical research on green GDP accounting, scholars are moving towards the direction of enrichment as a whole due to regional differences and different key calculation contents in different historical periods when accounting for green GDP. Based on the theory of green GDP, this paper summarizes the practical experience of predecessors, and adopts the following accounting formula according to the feasibility of data and the rationality of the algorithm:

Green GDP = traditional GDP - cost of resource consumption - cost of environmental degradation + benefit of resource and environment improvement

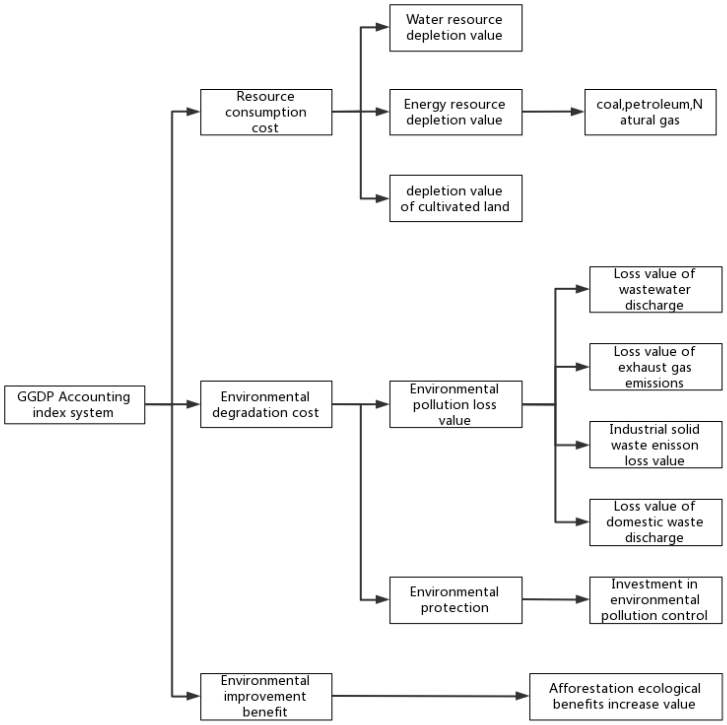


Figure 1. Green GDP accounting indicator system

3. Explore the correlation of each factor

3.1. Dimensionality reduction processing

Exploring the correlation between global temperature and various influencing factors is a problem with multiple inputs and single output. The relevant characteristic indicators can include time, location, air quality, pollutant types, rainfall, land index and so on up to 18, so it is difficult to directly explore their correlation relationship. Therefore, this paper uses the principal component analysis method to reduce the dimension of the data, and analyzes the statistical law of 18 characteristic indicators from the perspective of principal components. In order to ensure that "the amount of information is not lost", PCA technology reduces the dimension of the original features, that is, the original features are projected on the dimensions with the largest amount of projection information as far as possible. The original features are projected onto these dimensions to minimize the loss of information after dimensionality reduction. The formula for calculating the principal components is as follows:

$$z_1 = l_{11}x_1 + l_{12}x_2 + \cdots + l_{1p}x_p \tag{3}$$

$$z_2 = l_{21}x_1 + l_{22}x_2 + \cdots + l_{2p}x_p \tag{4}$$

Where z_1 and z_2 are the principal components, and l is the weight of each index in the principal components.Then, according to the weight of each characteristic index in different principal components, the statistical law of each index is analyzed.

3.2. The correlation between pairs is calculated

The correlation between pairs is calculated. It is calculated as follows.

$$r_s = 1 - \frac{6\sum d_i^2}{n^3 - n} \tag{5}$$

Where $d_i=x_i-y_i$, x_i and y_i are the rank of the two variables sorted by size (or goodness) respectively and n is the size of the sample. So n is equal to 2425.

Then use CART4.0 decision trees for regression: starting from the root node, test a certain feature of the instance, assign the instance to its child nodes according to the test results, at this time each child node corresponds to a value of the feature, so recursively test and assign the instance until it reaches the leaf node, and finally divide the instance into the class of the leaf node. Finally, the XGBoost lift tree is established, and the expression of the Xgboost objective function can be derived from the correlation loss, as shown in the following equation

$$\xi^{(t)}(q) = -\frac{1}{2}\sum_{j=1}^{T}\frac{(\sum g_i)^2}{\sum h_i + \lambda} + \gamma T \tag{6}$$

Finally, the classification tree can be obtained and the correlation analysis can be carried out by only calculating according to the maximum entropy increase and excluding the importance order.

3.3. Solving the Model

Due to the large number of characteristic indicators, this paper selects the nine largest coefficients, so as to reduce the dimension of the indicators to ten dimensions. Relevant indicators are: time, latitude, longitude, carbon dioxide content, methane content, sulfur dioxide, nitric oxide, QPI air index, land index.

Table 3. Related indicators

Common factor variance			
Indicators	Extract	Indicators	Extract
Time	0.575	RAVLT_immediate	0.710
Latitude	0.552	RAVLT_learning	0.488
Longitude	0.295	RAVLT_forgetting	0.928
CO2	0.174	RAVLT_pere_forgetting	0.928
CH4	0.253	LDELTOTAL	0.735
SO2	0.547	TRABSCOR	0.542
NO	0.731	FAQ	0.634
QPI	0.822	mPACCdigit	0.928
Land index	0.760	mPACCtrailsB	0.950

Then by drawing the Spearman correlation coefficient between the characteristic index and the global temperature index, as shown in the figure:



Figure 2. Characteristic variable heat map

As can be seen from the above figure, the global temperature has a strong correlation based on the results of the baseline period, namely QPI, carbon dioxide and location. It is positively correlated with time, indicating that the global temperature has a rising trend with the extrapolation of time, but the correlation with sulfur dioxide and NO is very weak, which is not consistent with facts and references. Therefore, this paperfurther uses random forest and XGBoost boosting tree model decision tree model to explain the correlation between data characteristics and global temperature through feature importance.

4. The impact of GGDP on China

4.1. Model Building

At present, the widely used predictive modeling methods include feedforward neural network, random forest, support vector regression and extreme gradient boosting tree, etc. Therefore, this paper

uses two of them to model separately, and then selects the best model. The two models are described as follows:

Feedforward neural networks(FNN)

A feedforward neural network is a simple gradient descent method that aims to minimize the error in the calculated output of the network. Its training process is continuously adjusted between weights and thresholds in order to reduce the network error to a predetermined minimum or stop at a predetermined training step.

(2) Random forest(RF)

As a widely used ensemble model, random forest has been successfully applied in prediction problems. The improved random forest algorithm is adopted in this paper, and the improvement idea is as follows: using the repeated sampling method with replacement to sample the sample set, so as to randomly generate k training sets, and further generate the corresponding decision tree. When the i-th decision tree is trained, the variance of the distance between all the predicted points and the actual points in the decision tree is calculated δ2(j)，The weights for the i-th decision tree are:

$$w_r(i) = \frac{1/\delta^2(i)}{\sum_{i=1}^{T} 1/\delta^2(j)} \tag{6}$$

Thus, the predicted value of the algorithm is obtained as follows.

$$\hat{\mu}(x) = \sum_{i=1}^{k} w_r(i) Z_t \tag{7}$$

Type, z_i for the i-th predicted values of the decision treeRandom grid search and 10-fold cross validation were used to determine the number of forests, the maximum depth of the decision tree, the minimum number of samples required for internal nodes, and the minimum number of samples required for leaf nodes.

4.2. Solving the Model

We calculate the loss value of natural resource depletion through user cost. The calculation of natural resource depletion loss value caused by economic activities should include resource loss caused by coal, oil and natural gas consumption and forestry harvesting, water resource consumption, fishery resource consumption and Marine resource consumption. However, due to the limitation of data, only mineral resources such as coal, oil and natural gas, forestry resources, water resources and fishery resources are counted in this paper. As a result, the calculated green GDP will be on the high side and smaller than the real value of green GDP. However, this does not affect the original intention of this paper to accelerate the accounting of environmental resources, establish a green GDP accounting system, and ensure sustainable.

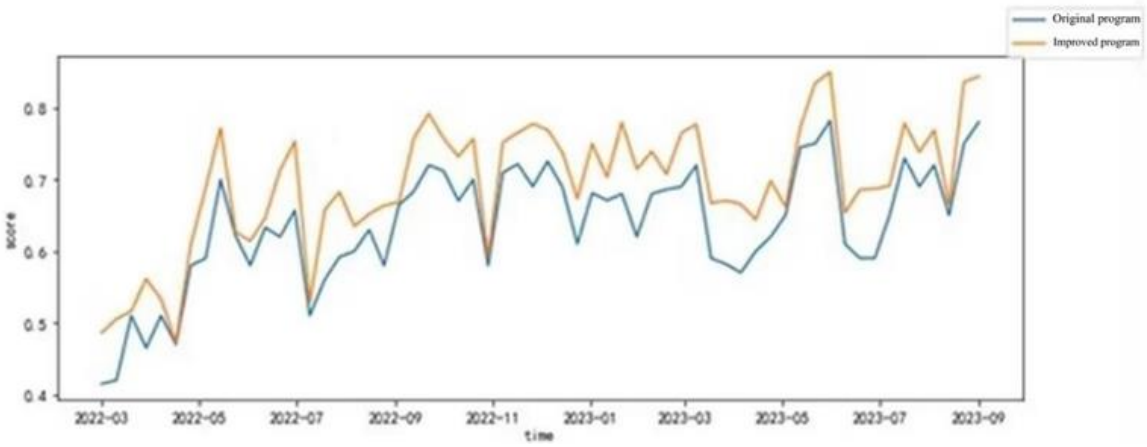


Figure 3. Depletion of natural resources due to economic activity Loss of value

5. Model evaluation

5.1. Strengths

Be able to properly model these factors and use this urea to establish a more scientific sustainable development plan.

When processing less eigenvalue data, the sample space of the data is not required to be large enough to solve the problem of less historical data, low sequence integrity and reliability, and more regular generated sequences can be obtained from irregular original data.

5.2. Weaknesses

Applicable to medium- and short-term forecast.

Do not take into account emergencies, war and other factors.

References

[1] Guo Hongyan,Fan Fengming, Study on environmental cost accounting of green GDP

[2] Wang Xiaoyan,Li Meizhou, Discussion on rank correlation coefficient and Spearman rank correlation coefficient

[3] Li Fanglin,Zhang Haibi.A comparative study of current GDP and green GDP

[4] Chen Yun,Dai Jinfang,Li Junjie ,CART decision tree classification method based on multiple features of images and its application.